DOCUMENT RESUME

ED 316 566 TM 014 517

AUTHOR Forehand, Garlie A.
TITLE Computer-Based Assessment for Remedial Instruction.
PUB DATE Aug 89
NOTE 8p.; Paper presented at the Annual Meeting of the American Psychological Association (97th, New Orleans, LA, August 11-14, 1989).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Students; Computer Assisted Instruction; *Computer Assisted Testing; *Educational Assessment; Evaluation Criteria; Higher Education; *Remedial Instruction; Standards

ABSTRACT

This paper focuses on computer-based assessment embedded in the process of instruction, with the assessment being used for placement, assignment, instructional feedback, progress assessment, and/or exit testing. This discussion is based on experience in developing and evaluating assessment and instruction materials for college-level remedial instruction. Such continuous measurement uses calibrated measures embedded in a curriculum to estimate, continuously and unobtrusively, dynamic changes in the student's proficiency. Criteria of product quality and competent interpretation for such measurement are reviewed. Three facets of assessment for instruction that pose new questions about standards and criteria are addressed: (1) specification and evaluation of "low-stakes" measurement that occurs during the course of instruction; (2) interpretation of measures taken during the course of learning, when proficiency is evolving within the process of instruction; and (3) communication with the learner as user of measurement information. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED316566

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
GARLIE A. FOREHAND
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Computer-Based Assessment for Remedial Instruction*

Garlie A. Forehand
Educational Testing Service

This paper will focus on computer-based assessment embedded in the process of instruction. Bunderson, Inouye, and Olson (1988) describe continuous measurement as the still-developing third generation of computerized educational measurement. Such measurement will "use calibrated measures embedded in a curriculum to continuously and unobtrusively estimate dynamic changes in the student's proficiency." What are the criteria of product quality and competent interpretation for such measurement?

This paper is based on experience in developing and evaluating assessment and instruction materials for college level remedial instruction (Forehand and Rice, 1988). This paper addresses three facets of assessment for instruction that pose new questions about standards and criteria.

- How to specify and evaluate "low-stakes" measurement that occurs during the course of instruction.
- How to interpret measures taken during the course of learning, when proficiency is taking new shapes in the process of instruction.
- How to communicate with the learner as user of measurement information.

* Presented at American Psychological Association, August 14, 1989. Part of a symposium on Computer Based Testing and Assessment chaired by Michael Kolen.

TM014517

The Stakes of Measurement

A student experiencing remedial instruction is likely to encounter assessment -- increasingly often computer-based -- used for five purposes.

> Placement, the determination that the student must take remedial courses, rather than regular freshman English or math.
>
> Assignment to an area and level of work, based on judgment about instructional strengths and needs.
>
> Instructional feedback, designed to help students learn from their own responses.
>
> Progress assessment, to communicate progress to learner and teacher.
>
> Exit testing, to determine when the student may leave the remedial track and enter regular courses.

These uses of assessment vary in the seriousness of their consequences -- the stakes. Closely correlated with the stakes is the time required to detect and reverse an unwarranted decision. Figure 1 displays these five uses of assessment in relation to reversibility. Figure 1 doesn't include the highest-stakes decisions, such as admission and certification, which may require years to reverse if they are reversible at all. The nice functional relationship is an accident; the horizontal axis is the order in which these tests are encountered in college remedial programs. The order may well be different in different applications. The time cost of an incorrect placement or exit decision must at least be measured in months -- the semester spent redundantly in the remedial course or unsuccessfully in non-remedial instruction. Inaccurate assignment or progress assessment can waste portions of a semester. The lowest stakes assessment provides opportunities to change course within minutes, by means of learner-friendly feedback, and

opportunities to confirm and clarify. Each kind of assessment requires its own standards and its own validation.

In assessment for feedback, the software is designed to play the role of a teacher giving feedback and responsive instruction; the role of professional judgment is played by decision rules and feedback messages. Measurement includes the processes of querying, scoring, deciding, delivering responsive feedback, summarizing across observations, and reporting. While risks may be low in comparison to selection and placement, risks are not zero. One can give erroneous instruction (e.g., reinforcing incorrect responses, encouraging dysfunctional habits, delivering inaccurate information); one can waste the learner's time; frustrate the learner with cul-de-sacs; and contribute to negative self-evaluation. How can the developer and user avoid these pitfalls? On the basis of our experience, we can suggest three principles: (a) build in the mechanisms for detecting and reversing wrong decisions; (b) design learner-interactions that help the learner understand outcomes and make some decisions; and (c) provide systematic opportunities for instructor review and override.

If low-stakes measurement is characterized by reversibility, then that reversibility must be built in. Decision rules need to include options that correct previous actions. Learners should branch out of instruction that is unneeded, branch into remediation of previously unrecognized faults, and be referred to instructors when the resolution is beyond the program's capacity. This implies that evaluation of the functioning of a system at a given moment always involves a sequence of events. The item as a familiar unit of psychometric analysis is replaced by the item-in-context. Context includes previous questions, subsequent questions, feedback, prior feedback, and

branching rules.

Computer-based assessment places great importance on the system's interaction with the learner. Opportunities to confirm a response before it is recorded will minimize the frustration of miskeying. Learner-friendly feedback is essential. The feedback should emphasize the learning process, not evaluation of the learner. Second tries enable the learner to experience success. The effectiveness of interaction must be engineered and evaluated.

Finally, a system of assessment for feedback needs a systematic opportunity for instructors to review and override the computer-controlled sequence. The software not only controls the presentation of test items, but also structures the learning situation. There are many opportunities for the student to get stuck, for the system to fail to move the student out of an unprofitable or discouraging sequence. The more intelligent the system becomes, the more a provision for human intervention is needed. As the routing system becomes more complex, there are increasing opportunities to encounter unforeseen loops and to accumulate misunderstandings of messages. Effective systems can vary from those that, relatively speaking, stand alone, to those that require frequent instructor input. Learners differ in their readiness to interpret and evaluate feedback messages. Therefore, instructor intervention may vary from occasional "hotline" help to repeated interaction. From the point of view of the developer and the user, the important consequence is that intervention opportunities must be designed and used.

Interpretation

The linear scale has been a valued tool of measurement theory. In

practice, measurement methodology assigns numbers to persons; a larger number implies more of the trait being measured than a lower number. In theory, it is postulated that the latent dimension lies along a scale, and that persons can be satisfactorily characterized by their positions on the scale. This model has worked well for many measurement applications. The concept of proficiency as a linear dimension, however, runs into limitations when one attempts to describe the status of a learner. Increase in ability is not the simple accumulation of new facts and skill. Learners reorganize their knowledge structures, automate procedures, chunk information to reorganize memory, develop strategies and models to tell them what is relevant and what is important.

As a model for the micro-level decisions required for assessment in instruction, the linear scale needs to be replaced with new models that incorporate cognitive and instructional theory as well as measurement principles. There has been much progress in this direction. Intelligent tutoring systems (e.g., Frederiksen and White, 1988) employ learner models at the most micro level of instruction. The individual's status is described by a model of the learner, a program that can be run to obtain a dynamic description the student at a given stage of progress. The student model is updated as learning proceeds, and compared with a model of expert performance. Many instructional problems do not lend themselves to such fine-grained modelling. For these kinds of situations, new theory is being developed that merges cognitive theory and measurement theory (Mislevy, in press, Embrettson, 1985, Tatsuoka, 1983). In this work, cognitive theory directs attention to instructionally relevant observations -- places in the process where measurement can affect instructional decisions. The task of measurement

theory is to specify how to make and summarize these observations in systematic, consistent, and valid ways.

The Learner as Interpreter

A new theme in the design and evaluation of assessment for instruction is the role of the learner as a user of assessment information. Standards emphasize responsibility of both designers and users, with the assumption that both are professionally accountable. Learners interpret and act on assessment results. Developers and users are in no position to tell learners how they must interpret feedback. The concept of the examinee as test interpreter raises a number of research and design issues.

- How do learners internally represent the learning situation -- the task, the requirements, the evaluation criteria, and their own learning process?
- What determines the learners' representations?
- How do different instructional strategies affect representations?
- How can the internal representations be accessed?
- What interventions are appropriate to prevent interactions that are unproductive or counterproductive?

Test standards must eventually be concerned with how learners do interpret feedback, and how the conditions of testing and teaching affect that interpretation.

These issues are currently matters for research. In the long term, they will influence ongoing development of test standards and guidelines. New models call for new evidence of construct validity. New ways of interpreting outcomes call for new modes of communication with examinees.

References

Bunderson, C. V., Inouye, D.K., and Olson, J.B. (1988) The Four Generations of Computerized Educational Measurement. In Linn , R. L. (Ed.) Educational Measurement, Third Edition. New York: Macmillan.

Forehand, G. A. and Rice, M. W. (1988) Diagnostic Assessment in Instruction. Machine Mediated Learning, 2, 287-296.

Embrettson, S. E. (1985) Multicomponent latent trait models for test design. In S.E. Embretson (Ed.), Test Design: Developments in Psychology and Psychometrics. Orlando, FL: Academic Press.

Mislevy, R. J. (In Press) Foundations of a New Test Theory. In N. Frederiksen, R. Mislevy, and I. Bejar (Eds.), Test Theory for a New Generation of Tests. Erlbaum.

Tatsuoka, K. K. (1983) Rule space: An approach for dealing with misconceptions based on item response theory. Journal of Educational Measurement, 20, 345-354.